



## LINEARLY SCALABLE FINITE IMPULSE RESPONSE FILTER

### PRIORITY CLAIM

[1] The present application claims priority from Indian Patent Application No. 1166/Del/2002 filed November 18, 2002, the disclosure of which is hereby incorporated by reference.

### 5 BACKGROUND OF THE INVENTION

[2] Technical Field of the Invention The present invention relates to Finite Impulse Response (FIR) Digital Filters. More specifically the invention relates to an efficient implementation for Finite Impulse Response filters in a multi-processor architecture.

Description of Related Art

[3] Finite Impulse Response (FIR) filters are an important and widely used form of a digital filter. FIR filters are used in numerous real time applications including telecommunication, particularly for the generation of constant phase delay characteristics. In addition, signal processing applications requiring an interpolation and decimation function incorporate FIR filtration as an integral part of the process.

[4] The output sample of a FIR filter is a convolution summation of input samples and the impulse response of the filter. The output,  $y(n)$  of a causal FIR filter can be written as:

$$y(n) = \sum_{k=0}^{H-1} h(k) * x(n-k)$$

Where:

H is the total number of filter coefficients and  $n = 0, 1, 2, 3...$  for different values of  $n$ , the output sample of the filter can be obtained;

$h(k)$  is the impulse response of the filter. Filter coefficients are determined for various values of  $k$ . The value of  $k$  can never be negative for a causal system; and

$x(m)$  is the input sample,  $m = n - k$  as shown in the equation above. The value of  $m$  can never be negative for a causal system.

As stated by the above equation, the coefficients are multiplied with the appropriate input samples and summed to obtain the output sample. The coefficients are multiplied with the appropriate input samples and then accumulated for obtaining a particular output sample. For  $N$  number of input samples and  $H$  number of coefficients, the required number of multiplications for a given output sample is  $H$ . The saturation point occurs at the  $H$ th output sample as shown in

FIGURE 1. The H number of multiplications are necessarily required if all the H coefficients are unique.

[5] Compared to other filters, FIR filters offer the following advantages: FIR filters are simple to implement and design. Linear-phase filters that delay the input signal, without causing phase distortion, can be easily realized using FIR filters. FIR filters do not employ any feedback and hence present fewer difficulties in practical implementation. The absence of feedback also simplifies design by making it possible to use finite precision arithmetic. Multi-rate applications such as “decimation” (reducing the sampling rate), “interpolation” (increasing the sampling rate), can be realized best by FIR filters.

[6] FIR filters can be easily implemented using fractional arithmetic unlike other types of filters in which it is difficult to do so. It is always possible to implement a FIR filter using coefficients with a magnitude of less than 1.0 as the overall gain of the FIR filter can be adjusted at its output. All the above advantages make FIR filters preferable for fixed-point Digital Signal Processors (DSPs).

[7] The conventional approach of FIR filter implementation utilizes a delay line ( $m = n - k$ , as in the equation above), resulting in increased memory requirements and slower computation.

[8] U.S. Patent No. 5,732,004 describes an algorithm for FIR filtering. It proposes a method of decimating and/or interpolating a multi-bit input signal in which  $n/2$  additions are performed, where ‘n’ is the number of bits in each filter coefficient. Scaling and multiplication of data with coefficients is performed using standard DSP architecture using coefficient values and associated scaling factors stored in memory. The coefficients are stored in coded form, and

are then decoded prior to multiplication by the data values. A delay line is essential for the implementation of this method.

[9] U.S. Patent Nos. 6,260,053 and 5,297,069 describe methods and implementations that utilize delay lines for FIR filter implementation. Moreover, none of these methods provides  
5 linear scalability.

#### SUMMARY OF THE INVENTION

[10] There is a need in the art to provide an efficient implementation for FIR filter for multi-processor architectures.

[11] There is also a need to provide a linearly scalable FIR filter.

10 [12] There is still further a need to obviate the need of a delay line for implementing an FIR filter thereby reducing the memory requirement for computation and resulting in a cost effective solution.

[13] There is also a need to speed up the computation of FIR filters.

[14] To address one or more of these needs, embodiments of the present invention are  
15 directed to an improved Finite Impulse Response (FIR) filter providing linear scalability and implementation without the need for delay lines. The FIR filter comprises a multiprocessor architecture including a plurality of ALUs (Arithmetic and Logic Unit), Multipliers units, Data cache, and Load/Store units sharing a common Instruction cache. A multi-port memory is also included. An assigning means assigns to each available processing unit the computation of  
20 specified unique partial product terms and the accumulation of each computed partial product on specified output sample values.

[15] The assigning means is a pre-process that is used to design the implementation of the FIR filter based on the filter specifications.

[16] Further the invention also provides a method for implementing an improved Finite Impulse Response (FIR) filter providing linear scalability using a multiprocessing architecture platform without the need for delay lines, comprising:

assigning to each available processing unit the computation of specified unique partial product terms; and

accumulating each computed partial product on specified output sample values.

#### BRIEF DESCRIPTION OF THE DRAWINGS

[17] A more complete understanding of the method and apparatus of the present invention may be acquired by reference to the following Detailed Description when taken in conjunction with the accompanying Drawings wherein:

[18] FIGURE 1 shows a generalized set of equations for the outputs of an FIR filter with N input/output samples, and H filter coefficients;

[19] FIGURE 2 shows a set of equations for the outputs of a FIR filter having 11 input/output samples, and 6 filter coefficients;

[20] FIGURE 3 shows a set of equations for the outputs of an FIR filter having 13 input/output samples, and 7 filter coefficients;

[21] FIGURE 4 shows a schematic diagram of a Multi-processor architecture; and

[22] FIGURE 5 shows a schematic diagram of the internal architecture of one processor of a multi-processor architecture.

## DETAILED DESCRIPTION OF THE DRAWINGS

[23] FIGURE 1 shows a generalized set of equations for the outputs of an FIR filter with N input/output samples, and H filter coefficients. These equations define the outputs from an FIR filter.

5 [24] FIGURE 2 refers to an example of the output samples of a symmetric FIR filter having 11 input/output samples, and 6 filter coefficients. If the total number of coefficients 'H' is an even number as shown in this example, then it is possible to obtain all the output samples by computing a subset of partial products that involve 'H/2' coefficients. Each of these partial products occurs in 2 symmetrically spaced samples, and can therefore be computed once and  
10 reused thereby eliminating the requirement of a delay line and reducing computational time and memory space requirements, as compared to the conventional approach of FIR filter implementation.

[25] In the example of FIGURE 2 output sample  $y(5)$  requires the computation of only 3 partial products, i.e.,  $x(5)h(0)$ ,  $x(4)h(1)$  and  $x(3)h(2)$ , as the other three partial product terms,  
15  $x(0)h(0)$ ,  $x(1)h(1)$  and  $x(2)h(2)$ , have already been computed in previous samples  $y(0)$ ,  $y(2)$  and  $y(4)$ , respectively. Similarly,  $x(5)h(0)$ ,  $x(4)h(1)$  and  $x(3)h(2)$ , contribute to terms  $y(10)$ ,  $y(8)$  and  $y(6)$ , respectively, and so on. It can be observed from FIGURE 2 that 'P' number of consecutive output calculations can be distributed over 'P' processors simultaneously, provided 'P' consecutive inputs are present. As shown for the case of two processors ( $P = 2$ ),  $y(5)$  and  $y(6)$   
20 can be calculated simultaneously.

[26] FIGURE 3 shows an example of the output samples of a symmetrical FIR filter having 13 input/output samples, and 7 filter coefficients. If the total number of coefficients 'H'

is an odd number as shown in this example, then it is possible to obtain all the output sample by calculating a maximum  $((H-1)/2+1)$  number of partial products after the saturation. When the filter coefficient are odd, i.e.,  $H$  is odd, the basic algorithm remains the same as for the even number of coefficients case with the exception for the  $((H-1)/2 + 1)$ th column around which the

5 FIGURE 3 is centro-symmetric. The total number of the required multiplications for each output sample is  $((H-1)/2 + 1)$  after the saturation. As shown, the product term associated with  $h(3)$  can not be reused to calculate the partial summation for any other output sample.

[27] The basic algorithm for symmetric FIR filter computation is as follows:

initialize loop index by zero;

10 load input sample and coefficients;

multiply input sample and coefficient;

update current output;

update partial output, if required (see Note below);

increment loop index by Number of processor; and

15 if loop termination is condition satisfied then stop, else go to second step.

Note: for Centro-symmetric case one product term is never re-used.

[28] For the case of a 2 parallel-processor architecture the basic algorithm is modified as follows:

initialize loop-index by zero;

20 load input sample for processor #1;

load coefficient for processor #1;

multiply input sample and coefficient in processor #1;

update current output in processor #1;

update partial output, if required, in processor #1;

load input sample for processor #2;

load coefficient for processor #2;

5 multiply input sample and coefficient in processor #2;

update current output in processor #2;

update partial output, if required, in processor #2;

increment loop-index by 2; and

if loop-index equals to  $(N*H)$  then stop, or go to second step.

10 The algorithm can be generalized for any number of processors in parallel.

[29] Further the algorithm is also applicable for the asymmetric FIR Filters in following steps:

initialize loop index by zero;

load input sample and coefficients;

15 multiply input sample and coefficient;

update current output;

increment loop index by Number of processor; and

if loop termination is condition satisfied then stop, else go to second step.

[30] Each processor in the parallel processing architecture is provided with an  
20 independent ALU (Arithmetic and Logic Unit), Multiplier unit, Data cache, and Load/Store unit.  
All the processors share a common Instruction cache, and multi-port memory. This architecture  
is typical of VLIW (Very Large Instruction Word) processors.



[31] FIGURE 4 shows the typical architecture of a Very Long Instruction Word (VLIW) processor core. It is a Multi-processor architecture in which each processor is composed of a mix of Register Banks, Constant Generators (immediate operands) and Functional Units. Different processors may have different unit/register mixes, but a single Program Counter and a unified I-cache controls them all, so that all processors run in lockstep (as expected in a VLIW). Similarly, the execution pipeline drives all processors. Inter-processor communication, achieved by explicit register-to-register move, is compiler-controlled and invisible to the programmer. At the multi-processor level, the typical architecture specification for such a device is as follows:

- an instruction delivery mechanism is provided to get instructions from the cache to the processors' data-path;
- an inter-processor communication mechanism is provided to transfer data among processors; and
- the data-cache is organized to establish a guarantee of main memory coherency in the presence of multiple memory accesses.

The Algorithm to compute output samples resides in the instruction fetch cache and expansion unit.

[32] FIGURE 5 shows the schematic diagram of the internal architecture of the processor. Generally a processor is a 4-issue (maximum 4 instructions can be issued simultaneously) VLIW core comprising the following:

four 32-bit integer ALUs, two 16x32 multipliers, one Load/Store Unit and one Branch Unit; and

sixty four 32-bit General-purpose registers and 8 1-bit branch registers (used to store branch condition, predicates and carries). Instructions allow two long immediate data operands per cycle.

[33] The Instruction Set Architecture is a very simple integer RISC instruction set with  
5 minimal “predication” support through select instructions.

[34] The memory-addressing repertoire includes base + offset addressing, allows speculative execution (dismissible loads, handled by the protection unit) and software pre-fetching.

[35] The instant invention utilizes this architecture to provide faster computation as it  
10 is evident from the fact that, if any FIR filter takes  $T$  units of time to compute  $N$  outputs in one processor, then for the increased number of processors (say  $P$ ), the time required to compute  $N$  outputs will be  $T/P$  units of time, which provides ‘Linear scalability’. Further, since all the processors load input samples and coefficients from the shared memory where both the data are available, there is no requirement for any delay line. The instant invention does not limit its  
15 scope to non-unique filter coefficients, but both the non-utilization of delay line as well as linear scalability can be realized for all unique coefficients.

[36] Although preferred embodiments of the method and apparatus of the present invention have been illustrated in the accompanying Drawings and described in the foregoing Detailed Description, it will be understood that the invention is not limited to the embodiments  
20 disclosed, but is capable of numerous rearrangements, modifications and substitutions without departing from the spirit of the invention as set forth and defined by the following claims.